

# The Distortion Rate Function of Cyclostationary Gaussian Processes

Alon Kipnis\*, Andrea J. Goldsmith\* and Yonina C. Eldar†

## Abstract

A general expression for the distortion rate function (DRF) of cyclostationary Gaussian processes in terms of their spectral properties is derived. This expression can be seen as the result of orthogonalization over the different components in the polyphase decomposition of the process. We use this expression to derive, in a closed form, the DRF of several cyclostationary processes arising in practice. We first consider the DRF of a combined sampling and source coding problem. It is known that the optimal coding strategy for this problem involves source coding applied to a signal with the same structure as one resulting from pulse amplitude modulation (PAM). Since a PAM-modulated signal is cyclostationary, our DRF expression can be used to solve for the minimal distortion in the combined sampling and source coding problem. We also analyze in more detail the DRF of a source with the same structure as a PAM-modulated signal, and show that it is obtained by reverse waterfilling over an expression that depends on the energy of the pulse and the baseband process modulated to obtain the PAM signal. This result is then used to study the information content of a PAM-modulated signal as a function of its symbol time relative to the bandwidth of the underlying baseband process. In addition, we also study the DRF of sources with an amplitude-modulation structure, and show that the DRF of a narrow-band Gaussian stationary process modulated by either a deterministic or a random phase sine-wave equals the DRF of the baseband process.

## I. INTRODUCTION

The distortion rate function (DRF) describes the average minimal distortion achievable in sending an information source over a rate-limited noiseless link. Sources with memory possess an inherent statistical dependency that can be exploited in the context of data compression. However, not many closed-form expressions for the DRF of such sources are known, and those are usually limited to the class of stationary processes. Two notable exceptions are the DRFs of the Wiener process, derived by Berger [1], and of auto-regressive Gaussian processes, derived by Gray [2]. Indeed, information sources are rarely stationary in practice, and source coding techniques that are based on stationary assumptions about the source will likely achieve poor performance if the source has time-varying statistics.

*Cyclostationary processes* (CS) (also known as *periodically correlated* processes or *block-stationary* processes) are a class of non-stationary processes whose statistics are invariant to time shifts by integer multiples of a given

\*Department of Electrical Engineering, Stanford University, CA

†Department of EE, Technion, Israel Institute of Technology, Haifa, Israel

This paper was presented in part at the 2014 international symposium on information theory, Honolulu, HI, USA.

time constant, denoted as the *period* of the process. As described in the survey by Gardner [3], CS processes have been used in many fields to model periodic time-variant phenomena. In particular, they arise naturally in synchronous communication where block coding and modulation by periodic signals are used. Spectral properties of CS processes, which will be used in our derivations, are also reviewed in [3] and in the references therein.

In this paper we analyze the DRF of CS Gaussian processes. Our study of the DRF of CS processes can be motivated by a very simple example: we are interested to find the DRF of the process obtained by modulating a continuous-time Gaussian stationary process  $U(\cdot)$  by a cosine wave with random phase  $\Phi$ , namely

$$X_\Phi(t) = \sqrt{2}U(t)\cos(2\pi f_0 t + \Phi), \quad t \in \mathbb{R}, \quad (1)$$

where  $\Phi$  is uniformly distributed over  $[0, 2\pi)$ . This process is commonly given as an example of a wide-sense stationary process in signal processing textbooks (e.g. [4, Ex. 8.18]). Note that due to the random phase,  $X_\Phi(\cdot)$  is not Gaussian and in fact is non-ergodic. It seems that in the context of rate-distortion theory, the spectrum of  $X_\Phi(\cdot)$  can only be used to derive an upper bound on its DRF given by the DRF of a Gaussian stationary process with the same second order statistics [5, Thm. 4.6.5]. The theory of asymptotic mean stationary (AMS) processes [6] implies that the DRF of  $X_\Phi(t)$  is given by the DRF of each one of its ergodic components [7, Thm. 11.3.1], corresponding to different values of the phase  $\varphi \in [0, 2\pi)$ . Each ergodic component satisfies a source coding theorem which allows us to evaluate its DRF using an optimization over probability distributions subject to a mutual information rate constraint. One might think this decomposition might provide us with a recipe to evaluate the DRF of  $X_\Phi(\cdot)$  by averaging over the DRF of the process  $X_\varphi(\cdot)$ , obtained by fixing the phase in  $X_\Phi(\cdot)$ . However, while the process  $X_\varphi(\cdot)$  is Gaussian, it is no longer stationary – but rather CS. While an expression for the DRF of CS processes is known [8], [7], it seems that the only existing mechanism for evaluating this DRF is by the Karhunen-Loève (KL) expansion [9]. In this method it is required to solve for the eigenvalues of a Fredholm integral equation for each finite blocklength, and use a waterfilling expression over these eigenvalues. The DRF is finally obtained in the limit as the size of the blocklength goes to infinity. This evaluation, however, does not exploit the special block periodicity of the CS source. Moreover, it does not provide any intuition on the optimal source coding technique in terms of spectral properties of the source. In contrast, the DRF for a stationary Gaussian process is obtained by waterfilling over its power spectral density, which provides clear intuition about how the source code represents each frequency component of the signal [10].

In this work we derive an expression for the DRF of Gaussian CS processes which uses their spectral properties, and therefore generalizes the waterfilling expression for the DRF of Gaussian stationary processes derived by Pinsker [11]. This expression is obtained by considering the polyphase components of the process, which can be seen as a set of stationary processes that comprise the CS process [12]. We show that the DRF of a discrete-time CS can be obtained in a closed form by orthogonalizing over these components at each frequency band. For continuous-time CS processes, we obtain an expression which is based on increasingly fine discrete-time approximations of the continuous-time signal. The DRF evaluated for these approximations converges to the DRF of the continuous-time

process under mild conditions on its covariance function.

The main results of this paper are divided into two parts. In the first part we derive a general expression for evaluating the DRF of a second order Gaussian CS process in terms of its spectral properties. This expression is given in the form of a reverse waterfilling solution over the eigenvalues of a spectral density matrix defined in terms of the *time-varying spectral density* of the source. For discrete-time Gaussian processes, the size of this matrix equals the discrete period of the source. We extend our result to Gaussian CS processes in continuous-time by taking increasingly finer discrete time approximations. The resulting expression is a function of the eigenvalues of an infinite matrix. In addition, we derive a lower bound on the DRF which can be obtained without evaluation of the matrix eigenvalues. We show that this bound is tight when the polyphase components of the process are highly correlated.

In the second part of the paper we use our general DRF expression to study the distortion-rate performance in more specific cases. Specifically:

- We derive a closed form expression for the DRF of a process with a pulse-amplitude modulation (PAM) signal structure. We show how this expression can be used to derive the minimal distortion in estimating a stationary Gaussian process from a rate-limited version of its sub-Nyquist samples.
- We study the effect of the symbol time in PAM on the information content of the modulated signal at the output of the modulator.
- We evaluate in a closed form the DRF of a Gaussian stationary narrowband process modulated by a deterministic cosine wave. We show that the DRF of the modulated process equals that of the baseband stationary Gaussian process provided the latter is narrowband. We further conclude that the stationary, non-Gaussian and non-ergodic process given by (1) above has DRF identical to the DRF of the modulated process without the random phase. These two results imply that the DRF of the stationary non-Gaussian amplitude modulated process is strictly smaller than the DRF of a Gaussian stationary process with the same second order statistics.

The rest of this paper is organized as follows: in Section II we review concepts and notation from the theory of CS processes and rate distortion theory. Our main results are given in Section III, where we derive an expression for the DRF of a Gaussian CS process. In Section IV we derive a lower bound on this DRF. In Section V we explore applications of our main result in various special cases. Concluding remarks are provided in Section VI.

## II. DEFINITIONS AND NOTATIONS

### A. Cyclostationary Processes

Throughout the paper, we consider zero mean Gaussian processes in both discrete and continuous time. We use round brackets to denote a continuous time index and square brackets for a discrete time index, i.e.

$$X(\cdot) = \{X(t), t \in \mathbb{R}\},$$

and

$$X[\cdot] = \{X[n], n \in \mathbb{Z}\}.$$

Matrices and vectors are denoted by bold letters.

The statistics of a zero mean Gaussian process  $X(\cdot)$  is specified in terms of its autocorrelation function<sup>1</sup>

$$R_X(t, \tau) \triangleq \mathbb{E}[X(t + \tau)X(t)].$$

If in addition the autocorrelation function is periodic in  $t$  with a fundamental period  $T_0$ ,

$$R_X(t + T_0, \tau) = R_X(t, \tau),$$

then we say that  $X(\cdot)$  is a *cyclostationary process* or simply *cyclostationary*<sup>2</sup> [3], [14]. We also assume that  $R_X(t, \tau)$  is bounded and Riemann integrable on  $[0, T_0] \times \mathbb{R}$ , and therefore

$$\sigma_X^2 = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \mathbb{E}X(t)^2 dt = \frac{1}{T_0} \int_0^{T_0} R_X(t, 0) dt$$

is finite.

Suppose that  $R_X(t, \tau)$  has a convergent Fourier series representation in  $t$  for almost any  $\tau \in \mathbb{R}$ . Then the statistics of  $X(\cdot)$  is uniquely determined by the *cyclic autocorrelation* (CA) function:

$$\hat{R}_X^n(\tau) \triangleq \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} R_X(t, \tau) e^{-2\pi i n t / T_0} dt, \quad n \in \mathbb{Z}. \quad (2)$$

The Fourier transform of  $\hat{R}_X^n(\tau)$  with respect to  $\tau$  is denoted as the *cyclic power spectral density* (CPSD) function:

$$\hat{S}_X^n(f) = \int_{-\infty}^{\infty} \hat{R}_X^n(\tau) e^{-2\pi i f \tau} d\tau, \quad -\infty \leq f \leq \infty. \quad (3)$$

If  $\hat{S}_X^n(f)$  is identically zero for all  $n \neq 0$ , then  $R_X(t, \tau) = R_X(0, \tau)$  for all  $0 \leq t \leq T$  and the process  $X(\cdot)$  is stationary. In such a case  $S_X(f) \triangleq \hat{S}_X^0(f)$  is the *power spectral density* (PSD) function of  $X(\cdot)$ . The *time-varying power spectral density* (TPSD) function [14, Sec. 3.3] of  $X(\cdot)$  is defined by the Fourier transform of  $R_X(t, \tau)$  with respect to  $\tau$ , i.e.

$$S_X^t(f) \triangleq \int_{-\infty}^{\infty} R_X(t, \tau) e^{-2\pi i f \tau} d\tau. \quad (4)$$

The Fourier series representation implies that

$$S_X^t(f) = \sum_{n \in \mathbb{Z}} \hat{S}_X^n(f) e^{2\pi i n t / T_0}. \quad (5)$$

Associated with every cyclostationary process  $X(\cdot)$  with period  $T_0$  is a set of stationary discrete time processes  $X^t[\cdot]$ ,  $0 \leq t \leq T_0$ , defined by

$$X^t[n] = X(T_0 n + t), \quad n \in \mathbb{Z}. \quad (6)$$

<sup>1</sup> In [14] and in other references, the symmetric auto-correlation function

$$\bar{R}_X(t, \tau) \triangleq \mathbb{E}[X(t + \tau/2)X(t - \tau/2)] = R_X(t - \tau/2, \tau),$$

the corresponding CPSD  $\hat{\bar{S}}_X^n(f)$  and TPSD  $\bar{S}_X^t(f)$ , are used. The conversion between  $\hat{S}_X^n(f)$  and the symmetric CPSD is given by  $\hat{\bar{S}}_X^n(f) = \hat{S}_X^n(f - n/(2T_0))$ .

<sup>2</sup>It is customary to distinguish between wide-sense cyclostationarity which relates only to the second order statistics of the process, and strict-sense cyclostationarity which relates to the finite order statistics of the process [15, Ch. 10.4]. Both definitions coincide in the Gaussian case.

These processes are called the *polyphase components* (PC) of the cyclostationary process  $X(\cdot)$ . The cross-correlation function of  $X^{t_1}[\cdot]$  and  $X^{t_2}[\cdot]$  is given by

$$\begin{aligned} R_{X^{t_1}X^{t_2}}[n, k] &= \mathbb{E}[X[T_0(n+k) + t_1]X[T_0n + t_2]] \\ &= R_X(T_0n + t_2, T_0k + t_1 - t_2) \\ &= R_X(t_2, T_0k + t_1 - t_2). \end{aligned} \quad (7)$$

Since  $R_{X^{t_1}X^{t_2}}[n, k]$  depends only on  $k$ , this implies that  $X^{t_1}[\cdot]$  and  $X^{t_2}[\cdot]$  are jointly stationary. The PSD of  $X^t[\cdot]$  is given by

$$\begin{aligned} S_{X^t}(e^{2\pi i\phi}) &\triangleq \sum_{k \in \mathbb{Z}} R_{X^tX^t}[0, k] e^{-2\pi i\phi k} \\ &= \sum_{k \in \mathbb{Z}} R_X(t, T_0k) e^{-2\pi i\phi k}, \quad -\frac{1}{2} \leq \phi \leq \frac{1}{2}. \end{aligned} \quad (8)$$

Using the spectral properties of sampled processes, we can use (8) and (5) to connect the functions  $S_{X^t}(e^{2\pi i\phi})$  and the CPSD of  $X(\cdot)$  as follows:

$$\begin{aligned} S_{X^t}(e^{2\pi i\phi}) &= \frac{1}{T_0} \sum_{k \in \mathbb{Z}} S_X^t\left(\frac{\phi - k}{T_0}\right) \\ &= \frac{1}{T_0} \sum_{k \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} \hat{S}_X^n\left(\frac{\phi - k}{T_0}\right) e^{2\pi i n t / T_0}. \end{aligned}$$

More generally, for  $t_1, t_2 \in [0, T_0]$  we have

$$\begin{aligned} S_{X^{t_1}X^{t_2}}(e^{2\pi i\phi}) &= \sum_{k \in \mathbb{Z}} R_{X^{t_1}X^{t_2}}[0, k] e^{-2\pi i\phi k} \\ &= \frac{1}{T_0} \sum_{k \in \mathbb{Z}} S_X^{t_2}\left(\frac{\phi - k}{T_0}\right) e^{2\pi i(t_1 - t_2)\frac{\phi - k}{T_0}} \\ &= \frac{1}{T_0} \sum_{k \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} \hat{S}_X^m\left(\frac{\phi - k}{T_0}\right) e^{2\pi i\left(m\frac{t_2}{T_0} + \frac{t_1 - t_2}{T_0}(\phi - k)\right)}. \end{aligned} \quad (9)$$

We now turn to briefly describe the discrete-time counterpart of the CA, CPSD, TPSD and the polyphase components defined in (2), (3), (4) and (6), respectively.

A discrete time zero mean Gaussian process  $X[\cdot]$  is said to be CS with period  $M \in \mathbb{N}$  if its covariance function

$$R_X[n, k] = \mathbb{E}[X[n+k]X[n]]$$

is periodic in  $k$  with period  $M$ . For  $m = 0, \dots, M$ , the  $m^{\text{th}}$  cyclic autocorrelation (CA) function of  $X[\cdot]$  is defined as

$$\hat{R}_X^m[k] \triangleq \sum_{n=0}^{M-1} R_X[n, k] e^{-2\pi i n m / M}.$$

The  $m^{\text{th}}$  CPSD function is then given by

$$\hat{S}_X^m(e^{2\pi i\phi}) \triangleq \sum_{k \in \mathbb{Z}} \hat{R}_X^m[k] e^{-2\pi i\phi k},$$

and the discrete TPSD function is

$$S_X^n(e^{2\pi i\phi}) \triangleq \sum_{k \in \mathbb{Z}} R_X[n, k] e^{-2\pi i\phi k}.$$

Finally, we have the discrete time Fourier transform relation

$$S_X^n(e^{2\pi i\phi}) = \frac{1}{M} \sum_{m=0}^{M-1} \hat{S}_X^m(e^{2\pi i\phi}) e^{2\pi i\phi nm/M}.$$

The  $m$ -th stationary component  $\bar{X}^m[\cdot]$ ,  $0 \leq m \leq M-1$  of  $X[\cdot]$  is defined by

$$X^m[n] \triangleq X[Mn+m], \quad n \in \mathbb{Z}. \quad (10)$$

For  $0 \leq m, r, n \leq M-1$  and  $k \in \mathbb{Z}$  we have

$$\begin{aligned} R_{X^m X^r}[n, k] &= \mathbb{E}[X^m[n+k]X^r[n]] \\ &= \mathbb{E}[X[Mn+Mk+m]X[Mn+r]] \\ &= R_X[Mn+r, Mk+m-r] \\ &= R_X[r, Mk+m-r]. \end{aligned} \quad (11)$$

Using properties of multi-rate signal processing:

$$\begin{aligned} S_{X^m X^r}(e^{2\pi i\phi}) &= \sum_{k \in \mathbb{Z}} R_X[r, Mk+m-r] e^{-2\pi i k \phi} \\ &= \frac{1}{M} \sum_{n=0}^{M-1} S_X^r\left(e^{2\pi i \frac{\phi-n}{M}}\right) e^{2\pi i(m-r)\frac{\phi-n}{M}}. \end{aligned} \quad (12)$$

The discrete-time counterpart of (9) is then

$$S_{X^m X^r}(e^{2\pi i\phi}) = \frac{1}{M} \sum_{k=0}^{M-1} \sum_{n=0}^{M-1} \hat{S}_X^n\left(e^{2\pi i \frac{\phi-k}{M}}\right) e^{2\pi i \frac{nr+(m-r)(\phi-k)}{M}}. \quad (13)$$

The functions  $S_{X^m X^r}(e^{2\pi i\phi})$ ,  $0 \leq m, r \leq M-1$  define an  $M \times M$  matrix  $\mathbf{S}_X(e^{2\pi i\phi})$  with  $(m+1, r+1)^{\text{th}}$  entry  $S_{X^m X^r}(e^{2\pi i\phi})$ . This matrix completely determines the statistics of  $X[\cdot]$ , and can be seen as the PSD matrix associated with the stationary vector valued process  $\mathbf{X}^M[n]$  defined by the stationary components of  $X[\cdot]$ :

$$\mathbf{X}^M[n] \triangleq (X^0[n], \dots, X^{M-1}[n]), \quad n \in \mathbb{Z}. \quad (14)$$

We denote the autocorrelation matrix of  $\mathbf{X}^M[\cdot]$  as the PSD-PC matrix. Note that the  $(r+1, m+1)^{\text{th}}$  entry of the PSD-PC matrix is given by (11).

## B. Examples

We present two important modulation models which result in CS processes.

**Example 1** (amplitude modulation (AM)). *Given a Gaussian stationary process  $U(\cdot)$  with PSD  $S_U(f)$ , consider the process*

$$X_{AM}(t) = \sqrt{2}U(t) \cos(2\pi f_0 t + \phi),$$

where  $f_0 > 0$  and  $\varphi \in [0, 2\pi)$  are deterministic constants. This process is CS with period  $T_0 = f_0^{-1}$  and CPSD [16, Eq. 41]

$$\hat{S}_{AM}^m(f) = \frac{1}{2} \begin{cases} S_U(f + f_0) + S_U(f - f_0), & m = 0, \\ S_U(f \mp f_0)e^{\pm 2i\varphi}, & m \pm 2, \\ 0, & \text{otherwise.} \end{cases}$$

This leads to the TPSD

$$S_X^t(f) = \frac{1}{2} S_U(f + f_0)(1 + e^{-2(2\pi i f_0 t + \varphi)}) + \frac{1}{2} S_U(f - f_0)(1 + e^{2(2\pi i f_0 t + \varphi)}). \quad (15)$$

**Example 2** (pulse-amplitude modulation (PAM)). Consider a Gaussian stationary process  $U(\cdot)$  modulated by a deterministic signal  $p(t)$  as follows:

$$X_{PAM}(t) = \sum_{n \in \mathbb{N}} U(nT_0)p(t - nT_0). \quad (16)$$

This process is CS with period  $T_0$  and CPSD [16, Eq. 49]

$$\hat{S}_{PAM}^n(f) = \frac{1}{T_0} P(f) P^* \left( f - \frac{n}{T_0} \right) S_U(f), \quad n \in \mathbb{Z}, \quad (17)$$

where  $P(f)$  is the Fourier transform of  $p(t)$  and  $P^*(f)$  is its complex conjugate. If  $T_0$  is small enough such that the support of  $P(f)$  is contained within the interval  $\left(-\frac{1}{2T_0}, \frac{1}{2T_0}\right)$ , then  $\hat{S}_{PAM}^n(f) = 0$  for all  $n \neq 0$ , which implies that  $X_{PAM}(\cdot)$  is stationary.

### C. The Distortion-Rate Function

For a fixed  $T > 0$ , let  $X_T$  be the reduction of  $X(\cdot)$  to the interval  $[-T, T]$ . Define the distortion between two waveforms  $x(\cdot)$  and  $y(\cdot)$  over the interval  $[-T, T]$  by

$$d_T(x(\cdot), y(\cdot)) \triangleq \frac{1}{2T} \int_{-T}^T (x(t) - y(t))^2 dt. \quad (18)$$

We expand  $X_T$  by a Karhunen-Loève (KL) expansion [9, Ch 9.7] as

$$X_T(t) = \sum_{k=1}^{\infty} X_k f_k(t), \quad -T \leq t \leq T, \quad (19)$$

where  $\{f_k\}$  is a set of orthogonal functions over  $[-T, T]$  satisfying the Fredholm integral equation

$$\lambda_k f_k(t) = \frac{1}{2T} \int_{-T}^T K_X(t, s) f_k(s) ds, \quad t \in [-T, T], \quad (20)$$

with corresponding eigenvalues  $\{\lambda_k\}$ , and where

$$K_X(t, s) \triangleq \mathbb{E}X(t)X(s) = R_X(s, t - s).$$

Assuming a similar expansion as (19) to an arbitrary random waveform  $Y_T$ , we have

$$\mathbb{E}d_T(X_T, Y_T) = \frac{1}{2T} \int_{-T}^T \mathbb{E}(X(t) - Y(t))^2 dt = \sum_{n=-\infty}^{\infty} \mathbb{E}(X_n - Y_n)^2.$$

The mutual information between  $X(\cdot)$  and  $Y(\cdot)$  on the interval  $[-T, T]$  is defined by

$$I_T(X(\cdot), Y(\cdot)) \triangleq \frac{1}{2T} \lim_{N \rightarrow \infty} I(\mathbf{X}_{-N}^N; \mathbf{Y}_{-N}^N),$$

where  $\mathbf{X}_{-N}^N = (X_{-N}, \dots, X_N)$ ,  $\mathbf{Y}_{-N}^N = (Y_{-N}, \dots, Y_N)$  and the  $X_n$ s and  $Y_n$ s are the coefficients in the KL expansion of  $X(\cdot)$  and  $Y(\cdot)$ , respectively.

Denote by  $\mathcal{P}_T$  the set of joint probability distributions  $P_{X, \hat{X}}$  over the waveforms  $(X(\cdot), \hat{X}(\cdot))$ , such that the marginal of  $X(\cdot)$  agrees with the original distribution, and the average distortion  $\mathbb{E}d_T(X(\cdot), \hat{X}(\cdot))$  does not exceed  $D$ . The rate-distortion function (RDF) of  $X(\cdot)$  is defined by

$$R(D) = \lim_{T \rightarrow \infty} R_T(D),$$

where

$$R_T(D) = \inf_{\mathcal{P}_T} I_T(X(\cdot); \hat{X}(\cdot)),$$

and the infimum is over the set  $\mathcal{P}_T$ . It is well known that  $R(D)$  and  $R_T(D)$  are non-decreasing convex functions of  $D$  [5], and therefore continuous in  $D$  over any open interval. We define their inverse function as the distortion-rate functions  $D(R)$  and  $D_T(R)$ , respectively. We note that by its definition,  $D(R)$  is bounded from above by the average power of  $X(\cdot)$  over a single period:

$$\begin{aligned} \sigma_X^2 &\triangleq \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \mathbb{E}X^2(t)dt = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T R_X(t, 0)dt \\ &= \frac{1}{T_0} \int_0^{T_0} R_X(t, 0)dt = \hat{R}_X^0(0). \end{aligned}$$

For Gaussian processes, we have the following parametric representation for  $R_T(D)$  or  $D_T(R)$  [9, Eq. 9.7.41]

$$D_T(\theta) = \sum_{k=1}^{\infty} \min\{\theta, \lambda_k\} \tag{21a}$$

$$R_T(\theta) = \frac{1}{2} \sum_{k=1}^{\infty} \log^+(\lambda_k/\theta), \tag{21b}$$

where  $\log^+ x \triangleq \max\{\log x, 0\}$ .

In the discrete-time case the DRF is defined in a similar way as in the continuous-time setting described above by replacing the continuous-time index in (18), (19) and (20), and by changing integration to summation. Since the KL transform preserves norm and mutual information, this definition of the DRF in the discrete-time case is consistent with standard developments for the DRF of a discrete-time source with memory as in [5, Ch. 4.5.2]. Note that with these definitions, the continuous-time distortion is measured in MSE per time unit while the discrete-time distortion is measured in MSE per source symbol. Similarly, in continuous-time,  $R$  represents bitrate, i.e., the number of bits per time unit. In the discrete-time setting we use the notation  $\bar{R}$  to denote bits per source symbol.

Since the distribution of a zero-mean Gaussian CS process with period  $T_0$  is determined by its second moment  $R_X(t, \tau)$ , we observe that such processes are  $T_0$ -ergodic and therefore *block-ergodic* as defined in [8, Def. 1]. It follows that a source coding theorem that associates  $D(R)$  with the optimal MSE performance attainable in encoding  $X(\cdot)$  at rate  $R$  is obtained from the main result of [8]. Specifically for the discrete-time case, it is shown



in [6, Exc. 6.3.1] that CS processes belong to the class of asymptotic mean stationary process (AMS) [6], where a source coding theorem for AMS processes can be found in [17].

#### D. Problem Formulation: Evaluating the DRF

In the special case in which  $X(\cdot)$  is stationary, it is possible to obtain  $D(R)$  without explicitly solving the Fredholm equation (20) or evaluating the KL eigenvalues: in this case, the density of these eigenvalues converges to the PSD  $S_X(f)$  of  $X(\cdot)$ . This leads to the celebrated *reverse waterfilling* expression for the DRF of a stationary Gaussian process, originally derived by Pinsker [11]:

$$R(\theta) = \frac{1}{2} \int_{-\infty}^{\infty} \log^+ [S_X(f) / \theta] df. \quad (22a)$$

$$D(\theta) = \int_{-\infty}^{\infty} \min \{S_X(f), \theta\} d\phi. \quad (22b)$$

The discrete-time version of (22) is given by

$$\bar{R}(\theta) = \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ [S_X(e^{2\pi i\phi}) / \theta] d\phi. \quad (23a)$$

$$D(\theta) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \min \{S_X(e^{2\pi i\phi}), \theta\} d\phi. \quad (23b)$$

Equations (22) and (23) define the distortion as a function of the rate through a joint dependency on the water level parameter  $\theta$ .

We note that stationarity is not a necessary condition for the existence of a density function for the eigenvalues in the KL expansion. For example, such a density function is known for the Wiener process [1] which is a non-stationary process.

The main problem we consider in this paper is the evaluation of  $D(R)$  for a general CS Gaussian process. In principle, this evaluation can be obtained by computing the KL eigenvalues in (20) for each  $T$ , using (21) to obtain  $D_T(R)$  and finally taking the limit as  $T$  goes to infinity. For general CS processes, however, an easy way to describe the density of the KL eigenvalues is in general unknown. As a result, the evaluation of the DRF directly by the KL eigenvalues usually does not lead to a closed-form solution. In the next section we derive an alternative representation for the function  $D(R)$  which is based on an approximation of the kernel  $K_X(t, s)$  used in (20).

### III. MAIN RESULTS

In this section we derive our main results with respect to an expression for the DRF of a Gaussian CS which does not involve the solution of the Fredholm integral equation (20).

Our first observation is that in the discrete-time case, the DRF of a Gaussian CS process can be obtained by an expression for the DRF of a vector Gaussian stationary source. This expression is an extension of (23), which was derived in [18, Eq. (20) and (21)] and is given as follows:

$$D_{\mathbf{X}}(\theta) = \frac{1}{M} \sum_{m=1}^M \int_{-\frac{1}{2}}^{\frac{1}{2}} \min \{ \lambda_m(e^{2\pi i \phi}), \theta \} d\phi \quad (24a)$$

$$R(\theta) = \frac{1}{M} \sum_{m=1}^M \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{2} \log^+ [\lambda_m(e^{2\pi i \phi}) / \theta] d\phi, \quad (24b)$$

where  $\lambda_1(e^{2\pi i \phi}), \dots, \lambda_M(e^{2\pi i \phi})$  are the eigenvalues of the PSD matrix  $\mathbf{S}_{\mathbf{X}}(e^{2\pi i \phi})$  at frequency  $\phi$ . We have the following result:

**Theorem 1.** *Let  $X[\cdot]$  be a discrete-time Gaussian cyclostationary process with period  $M \in \mathbb{N}$ . The distortion rate function of  $X[\cdot]$  is given by*

$$D(\theta) = \frac{1}{M} \sum_{m=1}^M \int_{-\frac{1}{2}}^{\frac{1}{2}} \min \{ \lambda_m(e^{2\pi i \phi}), \theta \} d\phi \quad (25a)$$

$$\bar{R}(\theta) = \frac{1}{2M} \sum_{m=1}^M \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ [\lambda_m(e^{2\pi i \phi}) / \theta] d\phi, \quad (25b)$$

where  $\lambda_1(e^{2\pi i \phi}) \leq \dots \leq \lambda_M(e^{2\pi i \phi})$  are the eigenvalues of the PSD-PC matrix with  $(m+1, r+1)^{th}$  entry given by

$$\mathbf{S}_{X^m X^r}(e^{2\pi i \phi}) = \frac{1}{M} \sum_{n=0}^{M-1} S_X^r(e^{2\pi i \frac{\phi-n}{M}}) e^{2\pi i (m-r) \frac{\phi-n}{M}}. \quad (26)$$

*Proof:* A full proof can be found in Appendix A. The idea is to use the polyphase decomposition (12) and the stationary vector valued process  $\mathbf{X}^M[\cdot]$  defined in (14). The PSD matrix of the process is shown to coincide with the PSD-PC matrix of  $X[\cdot]$ . The proof shows that the DRF of  $X[\cdot]$  coincides with the DRF of  $\mathbf{X}^M[\cdot]$ . The result then follows by applying (24) to  $\mathbf{X}^M[\cdot]$ . ■

Equation (25) has the waterfilling interpretation illustrated in Fig. 1: the DRF is obtained by setting a single water-level over all eigenvalues of (26). These eigenvalues can be seen as the PSD of  $M$  independent processes obtained by the orthogonalization of the PC of  $X[\cdot]$ . Compared to the limit in the discrete-time version of the KL expansion (21), expression (25) exploits the CS structure of the process by using its spectral properties. These spectral properties capture information on the entire time-horizon and not only over a finite blocklength as in the KL expansion.

The following theorem explains how to extend the above evaluation to the continuous-time case.

**Theorem 2.** *Let  $X(\cdot)$  be a Gaussian cyclostationary process with period  $T_0$  and correlation function  $R_X(t, \tau)$  Lipschitz continuous in its second argument. For a given  $M \in \mathbb{N}$ , denote*

$$D_M(\theta_M) = \frac{1}{M} \sum_{m=1}^M \int_{-\frac{1}{2}}^{\frac{1}{2}} \min \{ \lambda_m(e^{2\pi i \phi}), \theta_M \} d\phi \quad (27a)$$

$$R(\theta_M) = \frac{1}{2T_0} \sum_{m=1}^M \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ [\lambda_m(e^{2\pi i \phi}) / \theta_M] d\phi, \quad (27b)$$

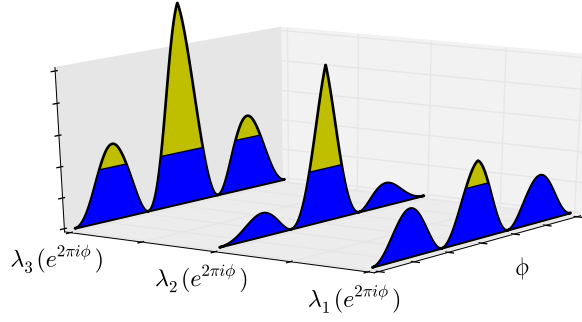


Fig. 1: Waterfilling interpretation of (25) for  $M = 3$ . The blue and the yellow parts are associated with equations (25a) and (25b), respectively.

where  $\lambda_1(e^{2\pi i\phi}) \leq \dots \leq \lambda_M(e^{2\pi i\phi})$  are the eigenvalues of the matrix  $\mathbf{S}_X(e^{2\pi i\phi})$  with its  $(m+1, r+1)^{th}$  entry given by

$$\begin{aligned} \frac{1}{T_0} \sum_{k \in \mathbb{Z}} S_X^{rT_0/M} \left( \frac{\phi - k}{T_0} \right) e^{2\pi i(m-r)\frac{\phi-k}{M}} \\ = \frac{1}{T_0} \sum_{k \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} \hat{S}_X^n \left( \frac{\phi - k}{T_0} \right) e^{2\pi i \frac{nr + (m-r)(\phi-k)}{M}}. \end{aligned} \quad (28)$$

Then the limit of  $D_M$  in  $M$  exists and the distortion-rate function of  $X(\cdot)$  is given by

$$D(R) = \lim_{M \rightarrow \infty} D_M(\theta_M(R)). \quad (29)$$

*Proof sketch:* The proof idea is to use a CS discrete-time process that approximates  $X(\cdot)$ . This approximation becomes tighter as  $M$  increases, so that the limit in (29) converges to the DRF of the continuous-time process coincides. The proof details are given in Appendix B.

### Discussion

The expression (27) is obtained by taking the limit in (25) over the time-period of a discrete-time CS process, where the code rate  $R$  is appropriately adjusted to bits per time unit. Although (27) only provides the DRF in terms of a limit, this limit is associated with the intra-cycle time resolution and not with the time horizon as in (21). This fact allows us to express the DRF in terms of the spectral properties of the process, which captures ‘memory’ in the process over the entire time horizon.

We note that limits of the form (29) have been obtained in closed-form using Szegő’s Toeplitz distribution theorem [19, Section 5.2] when the underlying process is stationary and the matrix considered is Toeplitz [20], [5], [2] or block Toeplitz [21], [22]. Unfortunately, the matrix in (28) is not Toeplitz or block Toeplitz so Szegő’s theorem is not applicable. In the following section we provide a few examples where the limit in (29) can be

obtained in closed form which lead to a closed form expression for the DRF.

Expression (27) can be seen as the extension to CS of the waterfilling expression (22) derived for stationary processes. While the latter can be understood as the limiting result of coding over orthogonal frequency bands [10], expression (27) implies that the DRF for CS processes is the result of two orthogonalization procedures: (1) over the PC inside a cycle, which is associated with the eigenvalues decomposition of the PSD-PC, and (2) over different frequency bands of the stationary processes resulting from the first orthogonalization.

The decomposition of the process into its stationary PCs can be further exploited to derive a lower bound on the DRF, which become tight as these PC become highly correlated. This bound is explored in the next section.

#### IV. LOWER BOUND

In this section we derive a lower bound on the DRF of a Gaussian CS process. This lower bound is expressed only in terms of the PCs of the process and does not require the eigenvalue evaluation of Theorems 1 and 2. The basis for this bound is the following proposition, which holds for any source distribution and distortion measure (although we will consider here only the quadratic Gaussian case).

**Proposition 1.** *Let  $\mathbf{X}[\cdot]$  be a vector valued process of dimension  $M$ . The distortion-rate function of  $\mathbf{X}[\cdot]$  satisfies*

$$D_{\mathbf{X}}(R) \geq \frac{1}{M} \sum_{m=0}^{M-1} D_{X_m}(R). \quad (30)$$

*Proof:* Any rate  $R$  code for the process  $\mathbf{X}[\cdot]$  induces a rate  $R$  code on each of the coordinates  $X_m[\cdot]$ ,  $m = 0, \dots, M-1$ . At each coordinate, this code cannot achieve lower distortion than the optimal rate  $R$  code for that coordinate. ■

Proposition 1 applied to Gaussian CS processes leads to the following result:

**Proposition 2.** *Let  $X[\cdot]$  be a discrete-time Gaussian CS process with period  $M \in \mathbb{N}$ . The distortion rate function of  $X[\cdot]$  satisfies*

$$D(\bar{R}) \geq \frac{1}{M} \sum_{m=0}^{M-1} \int_{-\frac{1}{2}}^{\frac{1}{2}} \min \{S_{X^m}(e^{2\pi i \phi}), \theta_m\} d\phi, \quad (31)$$

where for each  $m = 0, \dots, M-1$ ,  $\theta_m$  satisfies

$$\bar{R}(\theta_m) = \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ [S_{X^m}(e^{2\pi i \phi}) / \theta_m] d\phi. \quad (32)$$

Here

$$S_{X^m}(e^{2\pi i \phi}) \triangleq S_{X^m X^m}(e^{2\pi i \phi}) = \frac{1}{M} \sum_{n=0}^{M-1} S_X^m \left( e^{2\pi i \frac{\phi-n}{M}} \right)$$

is the PSD of the  $m^{\text{th}}$  PC of  $X[\cdot]$ .

*Proof:* The claim is a direct application of Proposition 1 to our case of a discrete-time CS process: the summands on the RHS of (31) are the individual DRF of the PCs  $X^m[\cdot]$ ,  $m = 0, \dots, M-1$ , of  $X[\cdot]$  obtained by (23).

Proposition 2 can be extended to the continuous-time case by approximating the outer integral in (33) by finite sums. This yields the following result:

**Proposition 3.** *Let  $X(\cdot)$  be a continuous-time Gaussian cyclostationary process with period  $T_0 > 0$  and correlation function  $R_X(t, \tau)$  Lipschitz continuous in its second argument. The distortion rate function of  $X(\cdot)$  satisfies*

$$D(R) \geq \frac{1}{T_0} \int_0^{T_0} \int_{-\frac{1}{2}}^{\frac{1}{2}} \min \left\{ \sum_{n \in \mathbb{Z}} S_X^t \left( \frac{\phi - n}{T_0} \right), \theta_t \right\} d\phi dt, \quad (33)$$

where for each  $0 \leq t \leq T_0$ ,  $\theta_t$  satisfies

$$R(\theta_t) = \frac{1}{2T_0} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ \left[ \sum_{n \in \mathbb{Z}} S_X^t \left( \frac{\phi - n}{T_0} \right) / \theta_t \right] d\phi. \quad (34)$$

*Proof:* See Appendix C.

The bound (31) is obtained by averaging the minimal distortion at rate  $R$  in describing each one of the PCs of  $X(\cdot)$ . For each such component  $X^t[\cdot]$  there is an associated water level  $\theta_t$  obtained by solving (34) for  $\theta_t$ . For  $R = 0$ ,  $\theta_t$  is always bigger than the essential supremum of

$$S_{X^t}(e^{2\pi i \phi}) = \frac{1}{T_0} \sum_{n \in \mathbb{Z}} S_X^t \left( \frac{\phi - n}{T_0} \right),$$

so the RHS of (31) equals the average over the total power of each one of the PCs of  $X(\cdot)$  which are summed to  $\sigma_X^2 = D_X(0)$ . On the other hand, if  $R \rightarrow \infty$  then  $\theta_t \rightarrow 0$  for all  $t \in [0, T_0]$ , and again equality holds in (31). That is, the bound is tight in the two extremes of  $R = 0$  and  $R \rightarrow \infty$ .

From a source coding point of view, the bound (33) can be understood as if a source code of rate  $R$  is applied to each of the PCs of  $X(\cdot)$  individually. On the other hand, the DRF in (27) is obtained by applying a single rate  $R$  code to describe all these PC simultaneously. As a result, the bound is tight only when all the PCs are maximally correlated, i.e. when a single PC determines the rest of them. A case where the latter hold is shown in the following example.

**Example 3** (equality in (33)). *Let  $X(\cdot)$  be the PAM signal of Example 2 where the pulse  $p(t)$  is given by*

$$p(t) = \begin{cases} 1 & 0 \leq t < T_0, \\ 0 & \text{otherwise.} \end{cases}$$

*The sample path of  $X(\cdot)$  has a staircase shape illustrated in Figure 2. This process is equivalent to the discrete time process  $\bar{U}[\cdot] \triangleq \{U(nT_0), n \in \mathbb{Z}\}$  both in information rate and squared norm per period  $T_0$ , which is enough to conclude that  $D_X(R) = D_U(RT_0)$ . Indeed, the PCs in this case are maximally correlated, in the sense that a realization of  $X^0[\cdot] = \{X(nT_0), n \in \mathbb{Z}\}$  determines the value of  $X^\Delta[\cdot] = \{X((n+\Delta)T_0), n \in \mathbb{Z}\}$  for all  $0 \leq \Delta < 1$ . In*

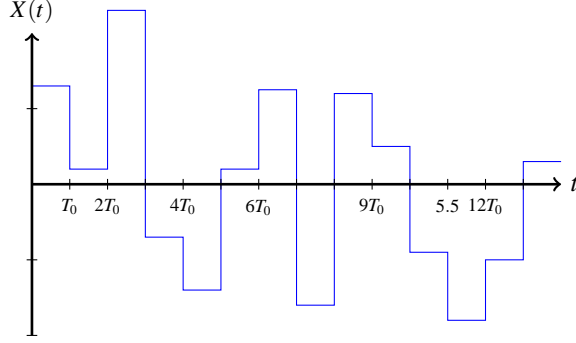


Fig. 2: An example of a continuous-time PAM process that attains equality in (33).

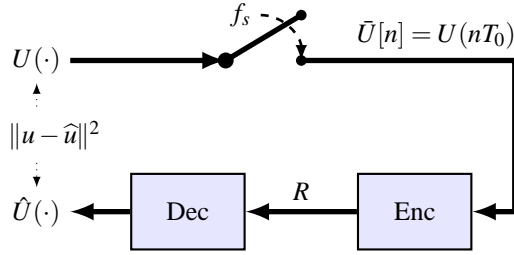


Fig. 3: Combined sampling and source coding system model.

addition, for all  $0 \leq t \leq T_0$  we have

$$S_X^t(e^{2\pi i\phi}) = S_X^0(e^{2\pi i\phi}) = \sum_{n \in \mathbb{Z}} S_U\left(\frac{\phi - n}{T_0}\right),$$

where the latter is the PSD of the discrete time process  $\bar{U}[\cdot]$ , so (22) implies that the RHS of (33) is the DRF of  $\bar{U}[\cdot]$ . We therefore conclude that the DRF of  $X(\cdot)$  is given by the RHS of (33).

## V. APPLICATIONS

In this section we apply the expression obtained in Theorem 2 to study the distortion-rate performance of a few CS processes that arise in practice.

### A. Combined Sampling and Source Coding

We begin with the distortion-rate performance in the combined sampling and source coding problem considered in [13]. This problem is described by the system of Figure 3: the source  $U(\cdot)$  is a Gaussian stationary process with a known PSD  $S_U(f)$ . The source is uniformly sampled at rate  $f_s = T_s^{-1}$ , resulting in the discrete time process  $\bar{U}[\cdot]$  defined by  $\bar{U}[n] = U(n/f_s)$ . The process  $\bar{U}[\cdot]$  is then encoded at rate  $R$  bits per time unit. The goal is to estimate the source  $U(\cdot)$  from its sampled and encoded version under a quadratic distortion. We denote by the function

$D_{U|\bar{U}}(f_s, R)$  the minimal distortion attainable in this estimation, where the minimization is over all collections of encoders and decoders operating at bitrate  $R$ . Note that if  $U(\cdot)$  is sampled above its Nyquist rate, then there is no loss of information in the sampling operation, and we get

$$D_{U|\bar{U}}(f_s, R) = D_U(R),$$

where  $D_U(R)$  is found by (22). Therefore, the case of most interest is that of sub-Nyquist sampling of  $U(\cdot)$ . In what follows we use Theorem 2 to derive  $D_{U|\bar{U}}(f_s, R)$  in closed form.

Our first observation is that the combined sampling and source coding problem of Fig. 3 can be seen as an indirect source coding problem [23]: the distortion is measured with respect to the process  $U(\cdot)$ , but a different process, namely  $\bar{U}[\cdot]$ , is available to the encoder. Wolf and Ziv [24] have shown that the optimal source coding scheme under quadratic distortion for this class of problems is obtained as follows: the encoder first obtains the minimal mean square error (MMSE) estimate of the unseen source, and then an optimal source code is applied to describe this estimated sequence to the decoder. In the setting of Figure 3, this implies that  $D_{U|\bar{U}}(f_s, R)$  is attained by first obtaining the MMSE estimate

$$\tilde{U}(t) = \mathbb{E}[U(t)|\bar{U}[\cdot]]$$

at the encoder, and then solving a standard source coding problem with the process  $\tilde{U}(\cdot)$  as the process to which the source code is applied. Moreover, this scheme implies that the distortion decomposes into 2 parts:

$$D_{U|\bar{U}}(f_s, R) = \text{mmse}(U|\bar{U}) + D_{\tilde{U}}(R), \quad (35)$$

where  $\text{mmse}(U|\bar{U})$  is the MMSE in estimating  $U(\cdot)$  from  $\bar{U}[\cdot]$ , and  $D_{\tilde{U}}(R)$  is the DRF of the process  $\tilde{U}(\cdot)$ .

Standard linear estimation techniques [25] leads to

$$\tilde{U}(t) = \sum_{n \in \mathbb{Z}} \bar{U}[n]w(t - nT_0) = \sum_{n \in \mathbb{Z}} U(nT_0)w(t - nT_0),$$

where the Fourier transform of  $w(t)$  given by

$$W(f) = \frac{S_U(f)}{\sum_{k \in \mathbb{Z}} S_U(f - k/T_0)}. \quad (36)$$

Moreover, the error in this estimation is

$$\text{mmse}(U|\bar{U}) = \int_{-\infty}^{\infty} S_U(f)df - \int_{-\frac{1}{2T_0}}^{\frac{1}{2T_0}} \tilde{S}_W(f)df, \quad (37)$$

where

$$\tilde{S}_W(f) = \sum_{k \in \mathbb{Z}} |W(f - k/T_0)|^2 S_U(f - k/T_0). \quad (38)$$

We conclude from the above that  $D_{\tilde{U}}(R)$ , and therefore  $D_{U|\bar{U}}(f_s, R)$ , is obtained by solving a source coding problem for an information source with a PAM signal structure, illustrated in Figure 4. Since Example 2 implies that such a signal is CS with period  $T_s = f_s^{-1}$ , we can apply Theorem 2 in order to evaluate this DRF. By doing so, we obtain the following general result:

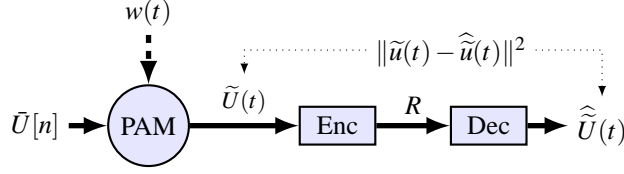


Fig. 4: Minimal distortion in Figure 3 is obtained by PAM followed by an optimal source code for the output of this modulation.

**Proposition 4** (DRF of PAM-modulated signals). *Let  $X_{PAM}(\cdot)$  be defined by*

$$X_{PAM}(t) = \sum_{n \in \mathbb{Z}} U(nT_0)p(t - nT_0), \quad t \in \mathbb{R}, \quad (39)$$

where  $U(\cdot)$  is a Gaussian stationary process with<sup>3</sup> PSD  $S_U(f)$  and  $p(t)$  is an analog deterministic signal with  $\int_{-\infty}^{\infty} |p(t)|^2 dt < \infty$  and Fourier transform  $P(f)$ . Assume moreover, that the covariance function  $R_{X_{PAM}}(t, \tau)$  of  $X_{PAM}(\cdot)$  is Lipschitz continuous in its second argument. The distortion-rate function of  $X_{PAM}(\cdot)$  is given by

$$D(\theta) = \frac{1}{T_0} \int_{-\frac{1}{2T_0}}^{\frac{1}{2T_0}} \min \{ \tilde{S}(f), \theta \} df \quad (40a)$$

$$R(\theta) = \frac{1}{2} \int_{-\frac{1}{2T_0}}^{\frac{1}{2T_0}} \log^+ [\tilde{S}(f)/\theta] df, \quad (40b)$$

where

$$\tilde{S}(f) \triangleq \sum_{k \in \mathbb{Z}} |P(f - k/T_0)|^2 S_U(f - k/T_0). \quad (41)$$

*Proof:* See Appendix D. ■

Proposition 4 applied to the process  $\tilde{U}(\cdot)$  implies that its DRF  $D_{\tilde{U}}(R)$  is given by waterfilling over the function

$$J(f) \triangleq \sum_{k \in \mathbb{Z}} |W(f - kf_s)|^2 S_U(f - kf_s).$$

As a result, we obtain from (35) and (40) the following expression for the minimal distortion in the combined sampling and source coding problem:

$$D_{U|\bar{U}}(f_s, R) = \text{mmse}(U|\bar{U}) + \frac{1}{T_0} \int_{-\frac{1}{2T_0}}^{\frac{1}{2T_0}} \min \{ J(f), \theta \} df, \quad (42a)$$

where

$$R(\theta) = \frac{1}{2} \int_{-\frac{1}{2T_0}}^{\frac{1}{2T_0}} \log^+ [J(f)/\theta] df. \quad (42b)$$

<sup>3</sup>Although we only use the value of  $U(t)$  at  $t \in \mathbb{Z}T_0$ , it is convenient to treat  $U(\cdot)$  as continuous-time source so that the expressions emerging have only continuous-time spectrum.



### B. Information Content of Signals with PAM Structure

Proposition 4 provides a general closed-form expression for the DRF of Gaussian processes with a PAM structure. In this subsection we use this expression to study the effect of the PAM of (39) on the distortion-rate curve of the signal  $X_{PAM}(\cdot)$  at the output of the modulator. Assuming that two processes have the same energy over time, the process with lower DRF can be described by fewer bits per second to the same distortion level. It is therefore intuitive to think about the DRF as a measure of the *information content* of the process<sup>4</sup>.

If we assume that the source for the symbols in the PAM is a Gaussian stationary process  $U(\cdot)$ , the output of the PAM of (39) can be seen as a non-ideal reconstruction of  $U(\cdot)$  from its uniform samples using pulses of shape  $p(t)$ , as illustrated in Figure 6. Since the randomness in  $X_{PAM}(\cdot)$  is only due to  $U(\cdot)$ , we expect  $X_{PAM}(\cdot)$  to have a smaller information content than  $U(\cdot)$ . In addition, we expect the information content of  $X_{PAM}(\cdot)$  to increase with the sampling rate  $1/T_0$ , and reach a saturation as this sampling rate exceeds the Nyquist rate of  $U(\cdot)$ . Indeed, when  $1/T_0$  is higher than the Nyquist rate of  $U(\cdot)$ , the support of  $S_U(f)$  is contained within  $\left(-\frac{1}{2T_0}, \frac{1}{2T_0}\right)$ . In this case, expression (40) implies that the DRF of  $X_{PAM}(\cdot)$  is obtained by waterfilling over the function

$$\tilde{S}(f) = |P(f)|^2 S_U(f). \quad (43)$$

That is, the effect of the modulation in super-Nyquist sampling is identical to the effect of a linear filter with frequency response  $P(f)$  applied to  $U(\cdot)$ . This filtering can be understood as a linear transformation of the coordinates [28, Ch. 22] represented by the frequency components. Assuming that  $P(f)$  does not change the support of (43) (that is, the change in ‘coordinates’ is invertible), the process  $U(\cdot)$  can be recovered from  $X_{PAM}(\cdot)$  with zero mean-square error. When the sampling frequency  $1/T_0$  goes below the Nyquist rate of  $U(\cdot)$ , perfect recovery of  $U(\cdot)$  is in general not possible. Intuitively, in this case we expect  $X_{PAM}(\cdot)$  to contain less information than  $U(\cdot)$ , and hence we should be able to describe it under the same normalized distortion level as  $U(\cdot)$  using fewer bits per time unit. A quantitative evaluation of this effect of the sampling rate is given in Figure 6, where the DRF of  $X_{PAM}(\cdot)$  is compared to the DRF of  $U(\cdot)$  for three sub-Nyquist sampling rates. Examples for the realization of  $X_{PAM}(\cdot)$  and  $U(\cdot)$  using sub- and super- Nyquist sampling rates are given in Figure 5.

### C. Amplitude Modulation with Random Phase

In this section we turn back to the two processes discussed in the introduction as our motivating examples and evaluate their DRFs using Theorem 2.

Consider the process  $X_{\Phi}(\cdot)$  obtained by modulating a stationary Gaussian process  $U(\cdot)$  by a cosine-wave of frequency  $f_0$  and a random phase  $\Phi$  uniform over  $[0, 2\pi)$ , as defined in (1). It is an elementary exercise [4, Ex. 8.18] to show that the process  $X_{\Phi}(\cdot)$  is stationary with PSD

$$S_{\Phi}(f) = \frac{1}{2} S_U(f - f_0) + \frac{1}{2} S_U(f + f_0). \quad (44)$$

<sup>4</sup>This notion is made precise by the notion of  $\epsilon$ -entropy [26], [27]

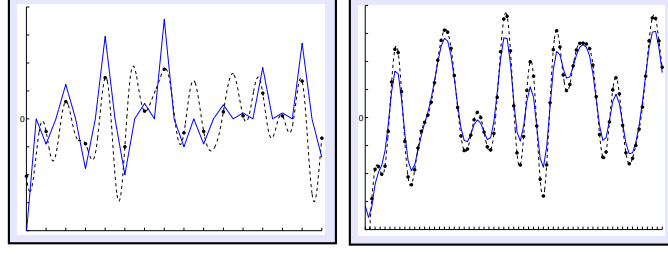


Fig. 5: Two realizations over time of the PAM process  $X_{PAM}(\cdot)$  (blue) and the baseband process  $U(\cdot)$  (dashed) with the PSD and pulse shape given in Figure 6, corresponding to sub-Nyquist (left) and super-Nyquist (right) sampling rates. Figure 6 below shows that for the same target distortion, the PAM-modulated process on the left is easier to describe than the PAM-modulated process on the right.

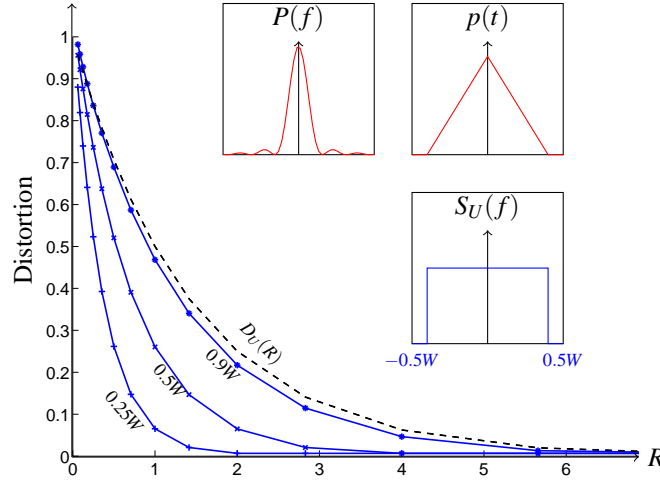


Fig. 6: The DRF of the PAM signal (16) for three values of sampling rate  $1/T_0$  compared to the Nyquist rate  $W$  of  $U(\cdot)$ . The DRF of the baseband stationary signal  $U(\cdot)$  (assuming an energy preserving modulation) is given by the dashed curve. The PSD of  $U(\cdot)$  and the shape of the pulse  $p(t)$  are given in the small frames. This figure shows that the information content of in the PAM process decreases with the sampling rate.

From [5, Thm. 4.6.5], an upper bound on the DRF of  $X_\Phi(t)$ , denoted by  $D_{X_\Phi}(R)$ , is obtained by the DRF of a Gaussian process with the same PSD  $S_\Phi(f)$  through the reverse-waterfilling (22). However, it seems that  $D_{X_\Phi}(R)$  cannot be determined solely from the second order statistics of  $X_\Phi(\cdot)$ .

The main obstacle in deriving  $D_{X_\Phi}(R)$  is the random phase of  $X_\Phi[\cdot]$ , which makes the process non-Gaussian and non-ergodic. This random phase can be handled using an asynchronous block code [7, Ch. 11.6], i.e. by adding a short prefix consisting of a source synchronization word to each block. Indeed, the following proposition follows directly from the proof of Theorem 11.6.1 in [7]:

**Proposition 5.** *For any  $\varphi \in [0, 2\pi)$  (deterministic), the DRF of the process  $X_\varphi(\cdot)$  coincides with the DRF of the process*

$$X_\varphi(t) = \sqrt{2}U(t)\cos(2\pi f_0 t + \varphi), \quad t \in \mathbb{R}. \quad (45)$$

It was noted in Example 1 above that  $X_\varphi(\cdot)$  is CS with the SCD function (15). It follows that  $D_{X_\varphi}(R)$  is given by the DRF of the Gaussian CS process  $X_\varphi(\cdot)$ , generated by modulating the stationary Gaussian process  $U(\cdot)$  using a deterministic cosine wave. Note that regardless of the carrier frequency  $f_0$ , the baseband process  $U(\cdot)$  can always be recovered from  $X_\varphi(\cdot)$ , and that the  $\sqrt{2}$  factor implies that the modulation preserves energy. These two facts are not enough to guarantee equality between the DRFs of the processes, since the modulation may lead to a ‘change in coordinates’ in the spectrum, in analogy with (43) and [28, Ch. 22]. In the following proposition we use Theorem 2 to show that this equality indeed holds as long as  $f_0$  is bigger than twice the bandwidth of  $S_U(f)$ .

**Proposition 6.** *Let  $U(\cdot)$  be a Gaussian stationary process bandlimited to  $(-f_B, f_B)$ . Let  $f_0 > 2f_B$ . The DRF of the process*

$$X_\varphi(t) = \sqrt{2}U(t)\cos(2\pi f_0 t + \varphi), \quad t \in \mathbb{R},$$

*equals the DRF of the stationary Gaussian process  $U(\cdot)$ .*

*Proof:* See Appendix E. ■

Proposition 7 asserts that the process  $X_\varphi(\cdot)$  with AM signal structure suffers the same minimal distortion as the baseband process  $U(\cdot)$  upon the encoding of each of them at rate  $R$ , and provided the latter is narrowband. Figure 7 shows that the above equality does not necessarily hold when  $U(\cdot)$  is not narrowband. Propositions 5 and 6 leads to the following conclusion:

**Corollary 7.** *Let  $U(\cdot)$  be a Gaussian stationary process bandlimited to  $(-f_B, f_B)$ . Assume that  $\Phi$  is uniformly distributed over  $(0, 2\pi)$  and  $f_0 > 2f_B$ . The distortion rate function of the stationary process*

$$X_\Phi(t) = \sqrt{2}U(t)\cos(2\pi f_0 t + \Phi), \quad t \in \mathbb{R},$$

*equals the DRF of the baseband process  $U(\cdot)$ .*

It is interesting to note that the DRF of a Gaussian process with the same PSD as the stationary process  $X_\Phi(\cdot)$  is strictly bigger than the DRF of the baseband process  $U(\cdot)$ , and therefore provides an upper bound to  $D_{X_\Phi}(R)$ . This upper bound is illustrated in Fig. 7.

## VI. CONCLUSIONS

We derived an expression for the distortion-rate function (DRF) of a class of Gaussian processes with periodically time-varying statistics, known as cyclostationary (CS) processes. This DRF is computed by reverse waterfilling over eigenvalues of a spectral density matrix associated with the polyphase components in the decomposition of the source. Unlike other general expression for the DRF of Gaussian processes that use orthogonal basis expansion over

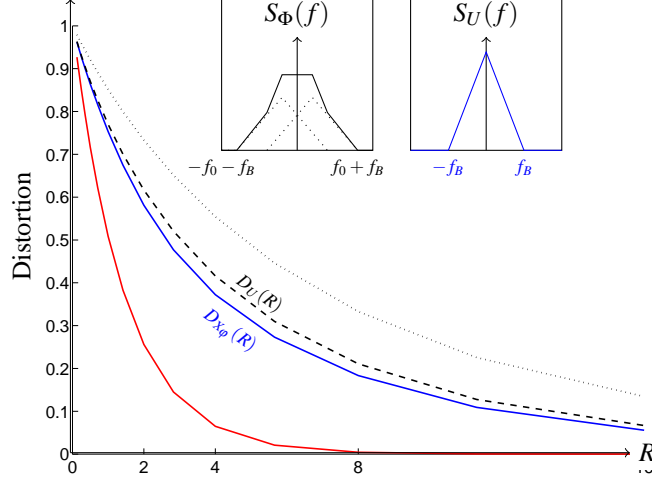


Fig. 7: The DRF of the processes  $X_\varphi(\cdot)$  of (45) (blue), the DRF of the baseband process  $U(\cdot)$  (dashed), and the lower bound of Proposition 3. The PSD  $S_U(f)$  is taken to be the pulse given in the small frame. Proposition 6 implies that  $D_U(R)$  and  $D_{X_\varphi}(R)$  coincides for  $f_s > 2f_B$ . Also shown is the DRF of the Gaussian stationary process with PSD  $S_\Phi(f)$  (dotted), which gives an upper bound to  $D_{X_\varphi}(R)$ .

increasing but finite time intervals, the expression we derive exploits the CS of the process by orthogonalizing the polyphase components. Since these components are defined over the entire time horizon, the resulting expression can be expressed in terms of the spectrum of the process. In the continuous-time counterpart the solution is given in terms of a limit over functions of these eigenvalues.

While we leave open the possibility whether there exists a closed form solution to the above limit in general, we have evaluated this limit in two special cases: a Gaussian CS processes with a PAM signal structure, and a Gaussian CS process with an amplitude modulation signal structure. As a result, we obtained the DRF of the processes obtained by these two important modulation schemes in terms of the power spectral density of the baseband stationary processes. We have also used the DRF result for a process with a PAM structure to derive the DRF of a process under combined sampling and source coding.

In addition to an expression for the DRF of CS processes, we have derived a lower bound on this DRF obtained by averaging the minimal distortion attained in encoding each of the polyphase components over a single period. This bound is tight when high correlation among these components is present.

## APPENDIX A

In this Appendix we provide a proof of Theorem 1. Consider the vector valued process  $\mathbf{X}^M[\cdot]$  defined in (14). The rate-distortion function of  $\mathbf{X}^M[\cdot]$  is given by (24):

$$D(\theta) = \frac{1}{M} \sum_{m=1}^M \int_{-\infty}^{\infty} \min \{ \lambda_m(e^{2\pi i \phi}), \theta \} d\phi, \quad (46a)$$

$$R(\theta) = \frac{1}{2} \sum_{m=1}^M \int_{-\infty}^{\infty} \log^+ [\lambda_m(e^{2\pi i \phi}) / \theta] d\phi, \quad (46b)$$

where  $0 \leq \lambda_1(e^{2\pi i \phi}) \leq \dots \leq \lambda_M(e^{2\pi i \phi})$  are the eigenvalues of the spectral density matrix  $\mathbf{S}_{\mathbf{X}^M}(e^{2\pi i \phi})$  obtained by taking the Fourier transform of covariance matrix  $\mathbf{R}_{\mathbf{X}}[k] = \mathbb{E}[X^M[n+k](X^M[n])^T]$  entry-wise. The  $(m, r)^{\text{th}}$  entry of  $\mathbf{S}_{\mathbf{X}^M}(e^{2\pi i \phi})$  is given by (12):

$$\begin{aligned} (\mathbf{S}_{\mathbf{X}^M}(e^{2\pi i \phi}))_{m,r} &= S_X^{m,r}(e^{2\pi i \phi}) \\ &= \frac{1}{M} \sum_{k=0}^{M-1} S_X^r(e^{2\pi i \frac{\phi-k}{M}}) e^{2\pi i (m-r) \frac{\phi-k}{M}}. \end{aligned} \quad (47)$$

It is left to show that the DRF of  $\mathbf{X}^M[\cdot]$  coincides with the DRF of  $X[\cdot]$ . By the source coding theorem for AMS processes [7, Thm. 11.4.1] it is enough to show that the operational block coding distortion-rate function ([7, Ch. 11.2]) of both processes is identical. Indeed, any  $N$  block codebook for  $\mathbf{X}^M[\cdot]$  is an  $MN$  block codebook for  $X[\cdot]$  which achieves the same quadratic distortion averaged over the block. However, since  $\mathbf{X}^M[\cdot]$  is stationary, by [7, Lemma. 11.2.3] we know that any distortion above the DRF of  $\mathbf{X}^M[\cdot]$  is attained for large enough  $N$ . This implies that the same is true for  $X[\cdot]$ .

## APPENDIX B

In this Appendix we prove Theorem 2. Given a Gaussian cyclostationary process  $X(\cdot)$  with period  $T_0 > 0$ , we define the discrete-time process  $\bar{X}[\cdot]$  obtained by uniformly sampling  $X(\cdot)$  at intervals  $T_0/M$ , i.e.

$$\bar{X}[n] = X(nT_0/M), \quad n \in \mathbb{Z}. \quad (48)$$

The autocorrelation function of  $\bar{X}[\cdot]$  satisfies

$$\begin{aligned} R_{\bar{X}}[n+M, k] &= \mathbb{E}[\bar{X}[n+M+k]\bar{X}[n+M]] \\ &= \mathbb{E}[X(nT_0/M + T_0 + kT_0/M)X(nT_0/M + T_0)] \\ &= R_X(nT_0/M + T_0, kT_0/M + T_0) \\ &= R_X(nT_0/M, kT_0/M) \\ &= R_{\bar{X}}[n, k], \end{aligned}$$

which means that  $\bar{X}[\cdot]$  is a discrete-time Gaussian cyclostationary process with period  $M$ . The TPSD of  $\bar{X}[\cdot]$  is given by

$$S_{\bar{X}}^m(e^{2\pi i \phi}) = \frac{M}{T_0} \sum_{k \in \mathbb{Z}} S_X^{mT_0/M} \left( \frac{M}{T_0}(\phi - k) \right).$$

This means that the PSD of the  $m^{\text{th}}$  PC of  $\bar{X}[\cdot]$  is

$$\begin{aligned} S_{\bar{X}}^m(e^{2\pi i\phi}) &= \frac{1}{M} \sum_{n=0}^{M-1} S_{\bar{X}}^m\left(e^{2\pi i\frac{\phi-n}{M}}\right) \\ &= \frac{1}{T_0} \sum_{n=0}^{M-1} \sum_{k \in \mathbb{Z}} S_X^{mT_0/M}\left(\frac{\phi - Mk - n}{T_0}\right) \\ &= \frac{1}{T_0} \sum_{l \in \mathbb{Z}} S_X^{mT_0/M}\left(\frac{\phi - l}{T_0}\right). \end{aligned}$$

By applying Theorem 1 to  $\bar{X}[\cdot]$ , we obtain an expression for the DRF of  $\bar{X}[\cdot]$  as a function of  $M$ :

$$D_M(\theta_M) = \frac{1}{M} \sum_{m=1}^M \int_{-\frac{1}{2}}^{\frac{1}{2}} \min\{\lambda_m(e^{2\pi i\phi}), \theta_M\} d\phi \quad (49a)$$

$$\bar{R}(\theta_M) = \frac{1}{2M} \sum_{m=1}^M \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ [\lambda_m(e^{2\pi i\phi}) / \theta_M] d\phi, \quad (49b)$$

where  $\lambda_1(e^{2\pi i\phi}) \leq \dots \leq \lambda_M(e^{2\pi i\phi})$  are the eigenvalues of the matrix with  $(m+1, r+1)^{\text{th}}$  entry

$$\begin{aligned} S_{\bar{X}^m \bar{X}^r}(e^{2\pi i\phi}) &= \frac{1}{M} \sum_{n=0}^{M-1} S_{\bar{X}}^r\left(e^{2\pi i\frac{\phi-n}{M}}\right) e^{2\pi i(m-r)\frac{\phi-n}{M}} \\ &= \frac{1}{T_0} \sum_{n=0}^{M-1} \sum_{k \in \mathbb{Z}} S_X^{rT_0/M}\left(\frac{\phi - n - kM}{T_0}\right) e^{2\pi i(m-r)\frac{\phi-n}{M}}, \\ &= \frac{1}{T_0} \sum_{l \in \mathbb{Z}} S_X^{rT_0/M}\left(\frac{\phi - l}{T_0}\right) e^{2\pi i(m-r)\frac{\phi-l}{M}}. \end{aligned} \quad (50)$$

In order to express the code-rate in bits per time unit, we multiply the number of bits per sample  $\bar{R}$  by the sampling rate  $M/T_0$ . This shows that the DRF of  $\bar{X}[\cdot]$ , as measured in bits per time unit  $R$ , is given by (27).

In order to complete the proof we rely on the following lemma:

**Lemma 8.** *Let  $X(\cdot)$  be as in Theorem 2 and let  $\bar{X}[\cdot]$  be its uniformly sampled version at rate  $M/T_0$  as in (48). Denote the DRF at rate  $R$  bits per time unit of the two processes by  $D(R)$  and  $\bar{D}(R)$ , respectively. Then*

$$\lim_{M \rightarrow \infty} \bar{D}(R) = D(R).$$

The rest of the appendix is devoted to the proof of Lemma 8.

Throughout the next steps it is convenient to use the covariance kernels  $K(t, s) = R_X(s, t - s)$  and  $\bar{K}[n, k] = R_{\bar{X}}[n, k - n]$ . For  $M \in \mathbb{N}$ , define

$$\tilde{K}(t, s) = K(\lfloor tM/T_0 \rfloor T_0/M, \lfloor sM/T_0 \rfloor T_0/M).$$

For any fixed  $T > 0$ , the kernel  $\tilde{K}(t, s)$  defines an Hermitian positive compact operator [29] on the space of square integrable functions over  $[-T, T]$ . The eigenvalues of this operator are given by the Fredholm integral equation

$$\tilde{\lambda}_l \tilde{f}_l(t) = \frac{1}{2T} \int_{-T}^T \tilde{K}(t, s) \tilde{f}_l(s) ds, \quad -T \leq t \leq T, \quad (51)$$

where it can be shown that there are at most  $MT/T_0$  non-zero eigenvalues  $\{\tilde{\lambda}_l\}$  that satisfy (51). We define the function  $\tilde{D}_T(R)$  by the following parametric expression:

$$\begin{aligned}\tilde{D}_T(\theta) &= \sum_{l=1}^{\infty} \min\{\tilde{\lambda}_l, \theta\} \\ R(\theta) &= \frac{1}{2} \sum_{l=1}^{\infty} \log^+ \left( \frac{\tilde{\lambda}_l}{\theta} \right)\end{aligned}\tag{52}$$

(the eigenvalues in (52) are implicitly depend on  $T$ ). Note that

$$\sum_{l=1}^{\infty} \tilde{\lambda}_l = \frac{1}{2T} \int_{-T}^T \tilde{K}(t, t) dt = \frac{1}{2T} \sum_{n=-N}^N K(nT_0/M, nT_0/M),\tag{53}$$

where  $N = MT/T_0$ . Expression (53) converges to

$$\frac{1}{2T} \int_{-T}^T K(t, t) dt \leq \sigma_X^2$$

as  $M$  goes to infinity due to our assumption that  $R(t, \tau)$  is Riemann integrable and therefore so is  $K(t, s)$ . Since we are interested in the asymptotic of large  $M$ , we can assume that (53) is bounded. This implies that  $\tilde{D}_T(R)$  is bounded.

We would like to claim that the eigenvalues  $\{\tilde{\lambda}_l\}$  approximate the eigenvalues  $\{\lambda_l\}$ . We have the following lemma:

**Lemma 9.** *Let  $\{\lambda_l\}$  and  $\{\tilde{\lambda}_l\}$  be the eigenvalues in the Fredholm integral equation of  $K(t, s)$  and  $\tilde{K}(t, s)$ , respectively. Assume that these eigenvalues are numbered in a descending order. Then*

$$|\lambda_l - \tilde{\lambda}_l| \leq 4CT_0/M, \quad l = 1, 2, \dots\tag{54}$$

*Proof of Lemma 9:* Approximations of the kind (54) can be obtained by Weyl's inequalities for singular values of operators defined by self-adjoint kernels [30]. In our case it suffices to use the following result [31, Cor. 1'']:

$$|\lambda_l - \tilde{\lambda}_l| \leq 2 \sup_{t, s \in [-T, T]} |K(t, s) - \tilde{K}(t, s)|, \quad l = 1, 2, \dots\tag{55}$$

The assumption that  $R_X(t, \tau)$  is Lipschitz continuous in  $\tau$  implies that there exists a constant  $C > 0$  such that for any  $t_1, t_2, s \in \mathbb{R}$ ,

$$|K(t_1, s) - K(t_2, s)| = |R_X(s, t_1 - s) - R_X(s, t_2 - s)| \leq C|t_1 - t_2|.$$

We therefore conclude that  $K_X(t, s)$  is Lipschitz continuous in both of its arguments from symmetry. Lipschitz continuity of  $K(t, s)$  implies

$$\begin{aligned}& |K(t_1, s_1) - K(t_2, s_2)| \\ & \leq |K(t_1, s_1) - K(t_1, s_2)| + |K(t_1, s_2) - \tilde{K}(t_2, s_2)| \\ & \leq C|s_1 - s_2| + C|t_1 - t_2|.\end{aligned}$$

As a result, (55) leads to

$$\begin{aligned}
|\lambda_l - \tilde{\lambda}_l| &\leq 2 \sup_{t,s} |K(t,s) - \tilde{K}(t,s)| \\
&= 2 \sup_{t,s \in [-T,T]} |K(t,s) - K(\lfloor tM/T_0 \rfloor T_0/M, \lfloor sM/T_0 \rfloor T_0/M)| \\
&\leq 2C(|t - \lfloor tM/T_0 \rfloor T_0/M| + |s - \lfloor sM/T_0 \rfloor T_0/M|) \\
&\leq 4CT_0/M,
\end{aligned}$$

which proves Lemma 9.

The significance of Lemma 9 is that the eigenvalues of the kernel  $K(t,s)$  used in the expression for the DRF of  $X(\cdot)$  can be approximated by the eigenvalues of  $\tilde{K}(t,s)$ , where the error in each of these approximations converge, uniformly in  $T$ , to zero as  $M$  increases. Since only a finite number of eigenvalues participate in (21) and since both  $D_T(R)$  and  $\tilde{D}_T(R)$  are bounded continuous functions of their eigenvalues, we conclude that  $\tilde{D}_T(R)$  converges to  $D_T(R)$  uniformly in  $T$ .

Now let  $\varepsilon > 0$  and fix  $M_0$  large enough such that for all  $M > M_0$  and for all  $T$

$$|D_T(R) - \tilde{D}_T(R)| \leq \varepsilon. \quad (56)$$

Recall that in addition to (23), the DRF of  $\tilde{X}[\cdot]$ , denoted here as  $\tilde{D}(\bar{R})$ , can also be obtained as the limit in  $N$  of the expression

$$\begin{aligned}
\bar{D}_N(\theta) &= \sum_{l=1}^{\infty} \min\{\tilde{\lambda}_l, \theta\} \\
\bar{R}(\theta) &= \frac{1}{2} \sum_{l=1}^{\infty} \log^+(\tilde{\lambda}_l/\theta),
\end{aligned}$$

where  $\tilde{\lambda}_1, \tilde{\lambda}_2, \dots$  are the eigenvalues in the KL expansion of  $\tilde{X}$  over  $n = -N, \dots, N$ :

$$\tilde{\lambda}_l f_l[n] = \frac{1}{2N+1} \sum_{k=-N}^N K_{\tilde{X}}[n,k] f_l[k], \quad l = 1, \dots, N, \quad (57)$$

(there are actually at most  $2N+1$  distinct non-zero eigenvalues that satisfies (57)). Letting  $T_N = T_0 M/N$  and  $\tilde{f}_l(t) = f_l(\lfloor t/T_0 \rfloor M)$  (57) can also be written as

$$\begin{aligned}
\tilde{\lambda}_l f_l[n] &= \int_{-T_N}^{T_N} \tilde{K}_X(nT_0/M, s) f_l(\lfloor s/T_0 \rfloor M) ds, \quad l = 1, 2, \dots, \\
\tilde{\lambda}_l \tilde{f}_l(t) &= \int_{-T_N}^{T_N} \tilde{K}(t, s) \tilde{f}_l(s) ds, \quad -T_N < t < T_N.
\end{aligned}$$

From the uniqueness of the KL expansion, we obtain that for any  $N$ , the eigenvalues of  $\tilde{K}(t,s)$  over  $T_N = T_0 M/N$  are given by the eigenvalues of  $\tilde{K}[n,k]$  over  $-N, \dots, N$ . We conclude that

$$\bar{D}_N(\bar{R}) = \tilde{D}_{T_N}(R), \quad (58)$$

where  $R = \bar{R}T_0/M$ . Now take  $N$  large enough such that

$$|\bar{D}_N(R) - \bar{D}(R)| < \varepsilon,$$



and

$$|D_{T_N}(R) - D(R)| < \varepsilon.$$

For all  $M \geq M_0$  we have

$$\begin{aligned} |D(R) - \bar{D}(R)| &= |D(R) - D_{T_N}(R) + D_{T_N}(R) + \tilde{D}_{T_N}(R) \\ &\quad - \tilde{D}_{T_N}(R) + \bar{D}_N(R) - \bar{D}_N(R) - \bar{D}(R)| \\ &\leq |D(R) - D_{T_N}(R)| \end{aligned} \tag{59}$$

$$+ |D_{T_N}(R) - \tilde{D}_{T_N}(R)| \tag{60}$$

$$+ |\tilde{D}_{T_N}(R) - \bar{D}_N(R)| \tag{61}$$

$$+ |\bar{D}_N(R) - \bar{D}(R)| \leq 3\varepsilon, \tag{62}$$

where the last transition is because: (59) and (62) are smaller than  $\varepsilon$  by the choice of  $N$ , (60) is smaller than  $\varepsilon$  from (56). and (61) equals zero from (58).

## APPENDIX C

In this Appendix we provide a proof of Proposition 3. We use the process  $\bar{X}[\cdot]$  defined in the proof of Theorem 2 as the uniform sampled version of  $X(\cdot)$  at rate  $T_0/M$ . From Proposition 1 we conclude that the DRF of  $\bar{X}[\cdot]$  satisfies

$$D_{\bar{X}}(\bar{R}) \geq \frac{1}{M} \sum_{m=0}^{M-1} \int_{-\frac{1}{2}}^{\frac{1}{2}} \min \left\{ \frac{1}{T_0} \sum_{l \in \mathbb{Z}} S_X^{mT_0/M} \left( \frac{\phi - l}{T_0} \right), \theta_m \right\} d\phi, \tag{63}$$

where for all  $m = 0, \dots, M-1$ ,  $\theta_m$  is determined by

$$\bar{R} = \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ \left[ \frac{1}{T_0} \sum_{l \in \mathbb{Z}} S_X^{mT_0/M} \left( \frac{\phi - l}{T_0} \right) / \theta_m \right] d\phi.$$

Denote  $t = mT_0/M$ . As  $M$  approaches infinity, the RHS of (63) converges to an integral with respect to  $t$  over the interval  $(0, T_0)$ , which implies

$$\bar{D}(\bar{R}) \geq \frac{1}{T_0} \int_0^{T_0} \int_{-\frac{1}{2}}^{\frac{1}{2}} \min \left\{ \sum_{l \in \mathbb{Z}} S_X^t \left( \frac{\phi - l}{T_0} \right), \theta_t \right\} d\phi, \tag{64}$$

and

$$\bar{R} = \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ \left[ \sum_{l \in \mathbb{Z}} S_X^{mT_0/M} \left( \frac{\phi - l}{T_0} \right) / \theta_m \right] d\phi, \tag{65}$$

where we denoted  $\theta_t = T_0 \theta_m$ . In order to go from  $\bar{R}$  to  $R$  we multiply (65) by  $M/T_0$ , so that (64) and (65) lead to (33). The fact that the function  $\bar{D}(R)$  converges to  $D(R)$  as  $M$  goes to infinity follows from the proof of Theorem. 2.

## APPENDIX D

In this Appendix we provide a proof of Proposition 4. The entries of the matrix  $\mathbf{S}(e^{2\pi i\phi})$  in Theorem 2 are obtained by using the CPSD of the PAM process (17) in (28). For all  $M \in \mathbb{N}$ , this leads to

$$\begin{aligned} \mathbf{S}_{m+1,r+1}(e^{2\pi i\phi}) &= \\ \frac{1}{T_0^2} \sum_{k \in \mathbb{Z}} &\left[ P\left(\frac{\phi-k}{T_0}\right) S_U\left(\frac{\phi-k}{T_0}\right) e^{2\pi i(\phi-k)\frac{m-r}{M}} \right. \\ &\times \left. \sum_{n \in \mathbb{Z}} P^*\left(\frac{\phi-n-k}{T_0}\right) e^{2\pi i\frac{nr}{M}} \right] \end{aligned} \quad (66)$$

$$\begin{aligned} &= \frac{1}{T_0^2} \sum_{k \in \mathbb{Z}} P\left(\frac{\phi-k}{T_0}\right) S_U\left(\frac{\phi-k}{T_0}\right) e^{2\pi i(\phi-k)\frac{m}{M}} \\ &\times \sum_{l \in \mathbb{Z}} P^*\left(\frac{\phi-l}{T_0}\right) e^{-2\pi i(\phi-l)\frac{r}{M}}. \end{aligned} \quad (67)$$

The expression (67) consist of the product of a term depending only on  $m$  and a term depending only on  $r$ . We conclude that the matrix  $\mathbf{S}(e^{2\pi i\phi})$  can be written as the outer product of two  $M$  dimensional vector, and thus it is of rank one. The single non-zero eigenvalue  $\lambda_M(e^{2\pi i\phi})$  of  $\mathbf{S}(e^{2\pi i\phi})$  is given by the trace of the matrix, which, by the orthogonality of the functions  $e^{2\pi i\frac{nr}{M}}$  in (66), is evaluated as

$$\lambda_M(e^{2\pi i\phi}) = \frac{M}{T_0^2} \sum_{k \in \mathbb{Z}} \left| P\left(\frac{\phi-k}{T_0}\right) \right|^2 S_U\left(\frac{\phi-k}{T_0}\right). \quad (68)$$

We now use (68) in (27). In order to obtain (40), we change the integration variable from  $\phi$  to  $f = \phi/T_0$  and the water-level parameter  $\theta$  to  $T_0\theta/M$ . Note that the final expression is independent of  $M$ , so the limit in (29) is already given by this expression.

## APPENDIX E

In this Appendix we provide a proof of Proposition 6. Since  $S_U(f)$  is compactly supported, the covariance function  $R_U(\tau) = \mathbb{E}U(t+\tau)U(t)$  is an analytic function and therefore Lipschitz continuous. Lipschitz continuity of  $R_U(\tau)$  implies Lipschitz continuity of  $R_X(t, \tau)$  in its second argument and therefore Theorem 2 applies: The DRF of the Gaussian CS process  $X_\phi(\cdot)$  with period  $T_0 = f_0^{-1}$  is obtained by using Theorem 2 with the SCD (15). For all  $M \in \mathbb{N}$  and  $m, r = 0, \dots, M-1$  we have,

$$\begin{aligned} \mathbf{S}_{m+1,r+1}(e^{2\pi i\phi}) &= \\ \frac{f_0}{2} \sum_{k \in \mathbb{Z}} &S_U(f_0(\phi-k-1)) \left(1 + e^{4\pi ir/M}\right) e^{2\pi i(\phi-k)\frac{m-r}{M}} \\ &+ \frac{f_0}{2} \sum_{k \in \mathbb{Z}} S_U(f_0(\phi-k+1)) \left(1 + e^{-4\pi ir/M}\right) e^{2\pi i(\phi-k)\frac{m-r}{M}}. \end{aligned}$$

Under the assumption that  $f_0 > 2f_B$  we have that for all  $\phi \in (-\frac{1}{2}, \frac{1}{2})$ ,  $S_U(f_0(\phi - k \pm 1)) = 0$  for all  $k \neq \pm 1$ . This leads to

$$\mathbf{S}_{m+1,r+1}(e^{2\pi i\phi}) \quad (69)$$

$$\begin{aligned} &= S_U(f_0\phi) \frac{f_0(1 + e^{4\pi i r/M})}{2} e^{2\pi i(\phi+1)\frac{m-r}{M}} \\ &\quad + S_U(f_0\phi) \frac{f_0(1 + e^{-4\pi i r/M})}{2} e^{2\pi i(\phi-1)\frac{m-r}{M}} \\ &= 2f_0 S_U(f_0\phi) e^{2\pi i\frac{m-r}{M}\phi} \cos\left(2\pi\frac{m}{M}\right) \cos\left(2\pi\frac{r}{M}\right). \end{aligned} \quad (70)$$

From (70) we conclude that the matrix  $\mathbf{S}(e^{2\pi i\phi})$  can be written as

$$\mathbf{S}(e^{2\pi i\phi}) = 2f_0 S_U(f_0\phi) \mathbf{S}_M(e^{2\pi i\phi}) \mathbf{S}_M^*(e^{2\pi i\phi}),$$

where  $\mathbf{S}_M(e^{2\pi i\phi}) \in \mathbf{R}^{M \times 1}$  is given by

$$\left(1, e^{2\pi i\phi/M} \cos\left(\frac{2\pi}{M}\right), \dots, e^{2\pi i\phi\frac{M-1}{M}} \cos\left(\frac{2\pi(M-1)}{M}\right)\right).$$

This means that  $\mathbf{S}(e^{2\pi i\phi})$  is a matrix of rank one, and its single non-zero eigenvalue is given by its trace:

$$\lambda_M(e^{2\pi i\phi}) = 2f_0 S_U(f_0\phi) \sum_{m=0}^{M-1} \cos^2(2\pi m/M) = M f_0 S_U(f_0\phi).$$

We use this in (28):

$$\begin{aligned} R_M(\theta) &= \frac{f_0}{2} \sum_{m=1}^M \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ [\lambda_m(e^{2\pi i\phi})] d\phi \\ &= \frac{f_0}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log^+ [M f_0 S_U(f_0\phi)/\theta] d\phi \\ &= \frac{1}{2} \int_{-\infty}^{\infty} \log^+ [S_U(f)/(\theta/M)] df, \end{aligned} \quad (71)$$

and

$$\begin{aligned} D_M(\theta) &= \frac{f_0}{M} \int_{-\frac{1}{2}}^{\frac{1}{2}} \min\{M f_0 S_U(f_0\phi), \theta\} d\phi \\ &= f_0 \int_{-\infty}^{\infty} \min\{S_U(f), \theta/M\} df \\ &= \int_{-\infty}^{\infty} \min\{S_U(f), \theta/M\} df. \end{aligned} \quad (72)$$

From (71) and (72) we conclude that for every  $M$ , the parametric expression of  $D$  as a function of  $R$  is identical to the DRF of the stationary process  $U(\cdot)$  given by (22).

#### ACKNOWLEDGMENT

This work was supported in part by the NSF under grant CCF-1320628, under the NSF Center for Science of Information (CSoI) grant CCF-0939370, and under BSF Transformative Science Grant 2010505. The authors wish to thank Robert M. Gray for helpful remarks. We also wish to thank to anonymous reviewers and the AE for the depth and details of their reviews which have greatly improved this paper.

## REFERENCES

- [1] T. Berger, "Information rates of Wiener processes," *IEEE Transactions on Information Theory*, vol. 16, no. 2, pp. 134–139, 1970.
- [2] R. Gray, "Information rates of autoregressive processes," *IEEE Transactions on Information Theory*, vol. 16, no. 4, pp. 412–421, 1970.
- [3] W. A. Gardner, A. Napolitano, and L. Paura, "Cyclostationarity: Half a century of research," *Signal Process.*, vol. 86, no. 4, pp. 639–697, Apr. 2006.
- [4] S. Haykin, *An Introduction to Analog and Digital Communications*. Wiley India Pvt. Limited, 2009.
- [5] T. Berger, *Rate-distortion theory: A mathematical basis for data compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
- [6] R. Gray, *Probability, Random Processes, and Ergodic Properties*. Springer, 2009.
- [7] —, *Entropy and information theory*. Springer-Verlag, 1990.
- [8] T. Berger, "Rate distortion theory for sources with abstract alphabets and memory," *Information and Control*, vol. 13, no. 3, pp. 254 – 273, 1968.
- [9] R. Gallager, *Information theory and reliable communication*. Wiley, 1968.
- [10] T. Berger and J. Gibson, "Lossy source coding," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2693–2723, 1998.
- [11] A. Kolmogorov, "On the shannon theory of information transmission in the case of continuous signals," *IRE Transactions on Information Theory*, vol. 2, no. 4, pp. 102–108, December 1956.
- [12] M. Vetterli, J. Kovačević, and V. K. Goyal, *Foundations of signal processing*. Cambridge University Press, 2014.
- [13] A. Kipnis, A. J. Goldsmith, Y. C. Eldar, and T. Weissman, "Distortion rate function of sub-Nyquist sampled Gaussian sources," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 401–429, Jan 2016.
- [14] W. Gardner, *Cyclostationarity in communications and signal processing*, ser. Electrical engineering, communications and signal processing. IEEE Press, 1994.
- [15] A. Papoulis and S. Pillai, *Probability, Random Variables and Stochastic Processes*, ser. McGraw-Hill electrical and electronic engineering series. McGraw-Hill Education, 2002.
- [16] W. Gardner, "Spectral correlation of modulated signals: Part i—analog modulation," *IEEE Transactions on Communications*, vol. 35, no. 6, pp. 584–594, 1987.
- [17] R. M. Gray, *Entropy and information theory*. Springer, 2011, vol. 1.
- [18] V. Kafedziski, "Rate distortion of stationary and nonstationary vector Gaussian sources," in *Statistical Signal Processing, 2005 IEEE/SP 13th Workshop on*, 2005, pp. 1054–1059.
- [19] U. Grenander and G. Szegő, *Toeplitz Forms and Their Applications*, ser. California monographs in mathematical sciences. University of California Press, 1958.
- [20] R. Gallager, *Information theory and reliable communication*, ser. Courses and lectures - International Centre for Mechanical Sciences. Springer-Verlag, 1970.
- [21] P. Tilli, "Singular values and eigenvalues of non-hermitian block toeplitz matrices," *CALCOLO*, vol. 33, no. 1-2, pp. 59–69, 1996.
- [22] Y. Chen, Y. C. Eldar, and A. J. Goldsmith, "Shannon meets Nyquist: Capacity of sampled Gaussian channels," *IEEE Transactions on Information Theory*, vol. 59, no. 8, pp. 4889–4914, 2013.
- [23] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Transactions on Information Theory*, vol. 8, no. 5, pp. 293–304, 1962.
- [24] J. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Transactions on Information Theory*, vol. 16, no. 4, pp. 406–411, 1970.
- [25] M. Matthews, "On the linear minimum-mean-squared-error estimation of an undersampled wide-sense stationary random process," *IEEE Transactions on Signal Processing*, vol. 48, no. 1, pp. 272–275, 2000.
- [26] E. C. Posner, E. R. Rodemich, and H. Rumsey Jr, "Epsilon entropy of stochastic processes," *The Annals of Mathematical Statistics*, pp. 1000–1020, 1967.
- [27] A. Kolmogorov, "Certain asymptotic characteristics of completely bounded metric spaces," *Doklady Akademii Nauk SSSR*, vol. 108, no. 3, pp. 385–388, 1956.
- [28] C. E. Shannon, "A mathematical theory of communication," *Bell System Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.
- [29] H. König, *Eigenvalue Distribution of Compact Operators*, ser. Operator Theory: Advances and Applications. Birkhäuser Basel, 2013.

- [30] H. Weyl, "Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung)," *Mathematische Annalen*, vol. 71, no. 4, pp. 441–479, 1912.
- [31] H. Wielandt, "Error bounds for eigenvalues of symmetric integral equations," in *Proc. Sympos. Appl. Math*, vol. 6, 1956, pp. 261–282.